APPROACHES TO RESEARCH DONOR EVALUATIONS:

A REVIEW OF THE ISSUES

W.D. Daniels, A.M. Young

International Development Research Centre

Presented at the
Research Donors' Group Meeting

December 10-11, 1987
Canberra, Australia

November 1987

DONOR EVALUATION METHODOLOGY

Table of Contents

1. Introduction

Participants at the December 1986 meeting of the Publicly Funded Research Donors (PFRD) Group decided that evaluation is a promising area for collaboration and IDRC agreed to prepare an issues paper as the basis for further discussion.

2. History of Evaluation Activities

Evaluation* is still a relatively new area of study and most of the literature in the field is less than 20 years old. Interest seems to have gone in waves since the 1950s with concern about project effectiveness leading to an increase in evaluation activity, only to have it decline when there was disillusionment with its limitations.

Throughout the 1970s, emphasis gradually shifted from project to program evaluation with a concurrent shift towards "comprehensive evaluation studies". In the 1980s, there has been a move away from these to more focussed "evaluations for management" and concern for the "impact" or effect of programs on the development process.

Evaluation developed first in industrial countries where the emphasis was placed on accountability, particularly by governments. The other major purposes of evaluation have been seen as "corporate memory" and "information for improving decision-making".

The level of interest and activity in the industrial countries continues to be high. In Canada for example, around 500 evaluation studies have been carried out for the federal government since 1979 when the Office of the Comptroller General was created**. Most federal departments and crown agencies have established their own evaluation units.

---

* See Annex A for definitions of the main types of evaluation and comments on the criteria used to evaluate research projects.

** Source: Office of The Comptroller General of Canada, November 1987.

2.1 Evaluation Activities in Developing Countries

A 1965 survey of the United Nations system's projects showed that only 14% of the 70 Third World countries surveyed had conducted systematic evaluations of the projects and 55% of the countries had not undertaken any evaluations at all (Forde and Sohm 1985). The limited resources devoted to evaluation in developing countries is particularly notable if one considers the relatively greater emphasis these countries have put on planning relative to the industrial countries. Of course, there are exceptions. India, with the assistance of the Ford Foundation, has had an established Programme Evaluation Organization located in its Planning Commission since 1952.

There is evidence, however, of increasing interest in evaluation in LDCs. A recent UNDP publication, Directory of Central Evaluation Authorities (1984), shows a large number (154) of the Third World countries have a central unit for evaluating development programs although the extent to which they are active is unclear.

2.2 LDC Research Evaluation

There has been almost no documented information available on the use of evaluation in developing country research systems although a recent publication (IDRC 254, 1987) presented 13 case studies from a Singapore workshop on the use of evaluation in agricultural research in ten countries and regional research institutions. These case studies confirmed the general impression that the use of evaluation as a management tool is quite limited. In most countries, external donor agencies have initiated the majority of formal evaluations.

The national managers present at this meeting concluded that ex ante appraisal probably provides the highest payoff and that the quality of analysis at this level was very good in a few cases but grossly inadequate in most. Monitoring, in contrast, is the area where most effort is expended and much wasted. Despite the large amount of paper produced, there is little critical analysis and recommendations that may be made often do not flow upwards to national managers. Non-donor initiated ex post evaluations are rare, although growing and are required by law or mandate in some institutions.

It was also clear from this meeting that there is a growing interest by national managers in increasing evaluations, particularly for demonstrating to policy-makers and the general public the benefits of investing in research. Pressures to improve efficiency of research are increasing since real financial resources per scientist have been declining in a number of countries even though the total number of scientists is still increasing. The need for evaluation will grow as national research systems become larger with more complex inter-relationships such as, for example, those in agricultural research between research on commodities and on farming systems programs.

2.3 Donor Agency Evaluation

Relative to total resources, donor agencies have probably devoted as high a level of resources to evaluation as any other sector. The recent report on aid effectiveness of the World Bank/IMF Task Force on Concessional Flows (the Cassen Report) documented an impressive amount of evaluation material produced by aid agencies. A report of the Development Assistance Committee (DAC) of the OECD identified 9,000 evaluations carried out by DAC agencies. At a conservative estimate, these agencies have spent some US $500 million on evaluation!

The situation is less clear with those donor agencies supporting research. The limited evidence from the literature and our survey of agencies participating in this meeting suggests that the level of evaluation activity is increasing. The growing involvement of the World Bank in supporting primarily agriculture and education research has certainly had an effect as the World Bank has a policy that two percent of the value of all loans be used for monitoring and evaluation.

One of the most significant reviews carried out by donor agencies in recent years was the "impact" assessment costing more than one million dollars (US) carried out by the CGIAR (1985).

Most evaluations of research appear to be penultimate or ex post assessments with few having good baseline or monitoring information for indicators on which to make judgements.

The two most significant developments appear to be growing interest in using evaluations to measure development effects ("impact") and, at least for general development agencies, in trying to generalize lessons of evaluations (e.g. World Bank study by Warren Baum). There does not appear to be any comparable drawing of lessons from research evaluations.

3. Value and Limitations (What Evaluations Can Do for Us)

3.1 Value

It is clear that donors see evaluation as a useful tool. The difficulties and uncertainties in selecting the best development options, the distance of donor head offices from the action, the fact that many of their projects are carried out by other organizations and thus not under donor control, the sometimes weak management capability and continuing questions about ODA effectiveness ensure that it will continue to be important.

There are important benefits that can be achieved by assessment at each of the three levels of ex ante, monitoring and ex post assessment (see Annex A) especially if the three are linked together more effectively. Most resources have been directed at the ex ante and monitoring levels. Ex ante assessment is probably the most important in that research is an area where getting your bets right in the first place is critical. However, evaluations have consistently shown a number of weaknesses in project design such as not addressing weak research and management capacity and poor links with users that could be corrected if these lessons were fedback into project design.

Monitoring activities have been found to be useful in encouraging flexibility and enhancing the success rate of projects, but as the case studies in the Singapore publication demonstrate, monitoring can be extremely time-consuming and of little value if it is not designed to provide measurement of change or used critically by managers.

For both methodological and conceptual reasons, least effort has been devoted to ex post evaluations and assessment of the development effects of research. It is perhaps the most difficult and, at the same time, the most promising area because of the potential value of such evaluations to contribute to improved project design and greater utilization of research results.

3.2 Limitations

It is useful to bear in mind limitations in what evaluations can do to avoid unrealistic expectations and waste of resources. Some of these limitations, we believe, can be overcome by more donor cooperation.

Evaluations can be useful in telling us how to do things better: they are usually not very helpful in guiding us on **what** to do. This

requires other approaches including better market research, the development of science and technology indicators, and resource allocation analysis.

Even in those areas where evaluations can contribute, weaknesses in methodology reduce their value. Most methodologies are still subjective and qualitative (see Annex A) and even these are often not carefully applied. Scientists can be remarkably unscientific when it comes to evaluating. Probably the most serious weaknesses are found when evaluations attempt to measure the utilization of research results and the economic benefits arising therefrom. There are so many researchers who contribute, from different locations over time, in developing a new technology or process and so many exogenous variables such as government support policies, market, pricing, etc. that affect the level of utilization of a research product, that it becomes extremely difficult to isolate and determine the payoff from any research project. Thus, the CGIAR-sponsored study had considerable difficulty in measuring "impact". If donor agencies want better information on the payoff from investment in research, we're going to have to devote more effort to developing better measurement techniques and to sharing this experience.

A third and possibly most limiting constraint arises from the use of evaluation findings. It is difficult to get sufficient value from evaluation expenditures if they are used for only a single purpose, and evaluation results are not generalized to guide policy. One way to improve the cost/benefit ratio could be to reduce the scope of evaluations. In many cases, 75% of the information collected can be achieved with the first 25% of the effort. A second and more important method is to increase the number of users and push for maximum involvement of the primary user -- the manager. In our experience, the process can be as important as the product. Finally, it appears that we could greatly increase the value of evaluations if we emphasize multi-purpose use to answer general policy as well as project specific questions. This possibility is raised in section 6 on donor cooperation.

4. Approaches of PFRD Group

As discussed, there is a wide variety in the kind of evaluations carried out by donor agencies. Fortunately, the range of approaches amongst research donors is somewhat narrower. The following notes summarize responses to five issues we raised with the donor agencies at this meeting.

With the exception of GATE, which uses the logical framework approach applied in GTZ, most agencies do not have rigorously structured evaluation systems.

4.1 What are the links between evaluation activities and the planning and policy process?

The materials at hand provide little explicit information about organizational structure and links between evaluation and planning, but the connection is implicit in that all agencies emphasize that evaluations are carried out to improve decision-making relating to future activities.

In IDRC, both activities are combined through the Office of Planning and Evaluation. It would appear that a similar arrangement exists in SAREC through the Secretariat for Information and Investigations. In ACIAR, the Research Program Coordinator plays a central role in linking these two processes. Senior management committees in all agencies are the final arbiters of issues relating to both processes.

4.2 What are the different kinds of evaluations undertaken?

Although evaluation in this paper has generally been taken to mean ex post evaluations, it is worth noting that most agencies indicated that they put a lot of effort into ex ante appraisals.

ACIAR in particular places considerable emphasis on this. Indeed, one of the core tasks of the project review (ex post) teams is to assess whether the ex ante appraisals of potential impact and spillover effects were accurate.

Most agencies evaluate both projects and programs (or groupings of related projects). The frequency and depth of review varies. ACIAR incorporates a review component into the project budget from the beginning. In IDRC, a project completion report (PCR) is required for every project but a full evaluation is usually only carried out for individual projects when particular questions are raised or there are implications for Centre policy.

In depth program reviews tend to be done on a less regular basis. In some agencies, they appear to be planned every three to five years. In others, programs are reviewed on an ad hoc basis at the request of senior management or the governing board.

4.3 Are evaluation consultants external to the agency and are they drawn from developing countries?

For most PFRD members, the majority of evaluations are managed internally with the others carried out either by external consultants or by teams which include external consultants. All agencies indicated that they make a special effort to ensure that developing country perspectives remain dominant. This is done by using developing country researchers as advisors or by including them on the review team.

IDRC has attempted to strengthen evaluation capacity directly, treating evaluation as a program activity in its own right and funding case studies and workshops.

4.4 Who decides whether and when evaluations are done?

ACIAR builds in an evaluation component into every project. IDRC is currently in the process of developing a more comprehensive evaluation plan which will incorporate and build on divisional and Centre-wide evaluations over a four-year period.

The decision to do project or program evaluations generally rests with project or program management, whereas evaluations of broader policy issues are made by senior management or the governing body.

Evaluations in IFS arise out of requests from the Sponsoring Committee for information on various aspects of the ongoing activities.

4.5 How does project monitoring relate to the evaluation process?

With the exception of ACIAR, evaluations and monitoring activities are carried out as distinct though inter-related processes.

BOSTID program staff constantly monitor the work of the grantees. Before grants are given, staff travel to the site to review the proposal with the principal investigator and evaluate the research capability of the researcher and the institution. Once a grant is approved, investigators submit brief activity reports at half year intervals, and comprehensive progress reports annually. Coordination meetings in each of the grant areas are convened annually so that the principal investigators can report on research results and discuss common problems.

Monitoring in IDRC is normally carried out by the program officer responsible for a project. Monitoring and technical reports are used by the program officer to prepare project completion reports. Qualitative evaluation information is gathered as a matter of course during the monitoring process.

Monitoring is central to the review process in ACIAR. Special emphasis is placed on accountability and measuring research effectiveness because of the "sunset clause" in their act of incorporation. The research program coordinators who conduct end-of-project reviews play a central role in project monitoring and implementation.

5. Increasing the Benefits from Evaluation

We believe there are a number of steps that can be taken to improve the potential benefits from any given level of resources devoted to evaluation. Three key areas where we believe greater donor efforts would be useful are reviewed below.

5.1 Building National Capability

The agencies participating in this group have a better reputation than some of the larger development agencies in being sensitive to national program needs and encouraging the development of national capability. There are opportunities to enhance national capability significantly over time if we maintain this same sensitivity in the evaluation field. There is growing resentment in developing countries about the onerous demands of overlapping evaluations, often initiated without consultation or input from national managers and carried out primarily by industrial country evaluators. **National organizations have the primary responsibility for program management including evaluation** and it is in our interest to encourage more national evaluation including maximum interaction with the donor agency on the timing, objectives and outcomes of any externally commissioned evaluations. This means identifying evaluation requirements as early as possible; encouraging national recipients to have input into the terms of reference including adding additional questions of interest to them and giving them an opportunity to comment on the final draft. Evaluations may be more accurate and insightful if national scientists are involved.

Just as each donor agency has been able to draw on the work over time of other agencies in building scientific capability, we will benefit from each other's efforts to build evaluation expertise in developing countries. A number of agencies probably already have a pool of scientists who they feel have good evaluation expertise that others could draw on.

5.2 Quality and Scope

Despite the growing level of resources devoted to evaluation, there has been relatively little effort devoted to developing and testing improved methodologies. It might be argued that research organizations in the industrial countries should be carrying the ball on this issue but work by any of the research donor agencies could help all if shared.

However, there are probably even greater benefits to be gained in the medium term by clarifying what questions can be answered by evaluation. Some questions may be best addressed by improved strategic forecasting and market research (e.g. how to identify the most critical gaps and the comparative advantage of donor agencies in funding different research areas).

The scope of evaluations, particularly relating to the efficiency of the research process, could be expanded to give more general answers (e.g. how to improve institutional capacity, what kinds of research training are most effective and how to link researchers and users more effectively). Addressing these broader questions may mean broadening the concept of evaluation to include research on research in which various different approaches are tested and evaluated. It also requires greater effort to generalize and draw lessons from the growing number of evaluations.

5.3 Evaluating the Benefits from Research

Evaluating the benefits achieved from research programs is undoubtedly the area of most interest to both donor agencies and national research programs. This was certainly the consensus of the national research managers present at the Singapore workshop referred to earlier. There is a growing number of studies of economic returns to agricultural research, often showing rates of return far above the rates achieved from other investments. Such studies may have a certain political value in demonstrating payoff although their reliability is questionable because they do not account for other factors that affect increases in productivity. There is little evidence that such studies have helped inform agencies about where to place their future bets. They have also typically generated information only on economic returns, neglecting the micro and macro effects on income levels, employment, exports, environment and other development objectives.

Even more difficult than measuring the benefits from improving general commodity yields, is measuring the specific effects of donor investment in research when, as mentioned earlier, any research result is the product of research over time and often by many institutions in different countries.

IDRC recently completed a review of the development effects from 10 projects (IDRC 246, 1986) and found that the development effects were generally modest and generally in proportion to the limited resources involved. Publishing these results, however, is only one step in generating the kind of interest and education of the public we would like to achieve. We are now embarking on a series of "tracer" studies which will measure the effects of new technology on individuals and communities over a period of several years. In this way, we hope to advance our understanding of the role of research in the development process. We also believe that information at this more personal level will prove more appealing to the general public, although new means of reaching the public through films and press are needed as well.

6. Donor Cooperation

This review suggests there is clearly some scope for increasing donor cooperation in this area.

(1) Joint Evaluations:

Recommendation: Donor agencies could consult IDRIS database to determine if other agencies are active in the institution they wish to evaluate and check to see if other donor agencies are interested in participating in or showing evaluation results. This would best be done on a bilateral basis.

There have been cases where donor agencies have cooperated in evaluating projects or institutions they're both funding. SAREC and IDRC have jointly evaluated several institutions and IDRC has also carried out a number of joint evaluations with the Ford Foundation. Joint evaluations cut the costs to individual agencies and reduce the often considerable amount of time required by the recipient institutions to respond to evaluation requirements. They may also provide a useful opportunity for donor agencies to share ideas and perceptions.

A review of 440 projects from five PFRD agencies showed 16 institutions have multiple donor agencies with ongoing projects. Thus, the University of Nairobi has three projects from IFS, two each from GATE and BOSTID and one from IDRC, for a total of eight ongoing projects supported by these agencies.

(2) Share Evaluation Results:

Recommendation: Donor agencies could agree to exchange their evaluation studies for use on a confidential basis.

In some cases, donor agencies may be able to satisfy their evaluation requirements by using a previous evaluation carried out by another donor for there are numerous instances documented of duplicative evaluations.

There may be some sensitivity about sharing evaluation results at this time given the critical comments that may be included. This is probably becoming less of an issue as a DAC evaluation meeting confirmed that most governments are pushing for greater public accountability of ODA programs. The Canadian government, for instance, has recently announced that evaluation results will be regularly published.

If this recommendation is pursued, it may be possible, at a later date, to put evaluation results on a common database allowing common computer access to summary information.

(3) Drawing Evaluation Lessons:

Recommendation: This donor meeting consider whether it would be feasible and valuable to try to generalize policy and programming lessons from research evaluation.

DAC has attempted, apparently with little success, to draw sector specific development lessons from the approximately 9,000 evaluations carried out by member countries of the OECD. However, we have not yet investigated this experience in detail. It might be easier to generalize in the more restricted field of development research but there would undoubtedly be difficulties even here, given the diversity of programs. IDRC has developed a database on some 40 policy and programming issues where we are interested in generalized results. However, even when we finish inputting the approximately 60 existing IDRC evaluations in 1988, we still expect the database to be too small.

REFERENCES

1) Evaluation in National Agricultural Research: Proceedings of a Workshop Held in Singapore, July 1986. 1987. Daniels, W.D. (ed.). International Development Research Centre, Ottawa, Ontario, Canada. IDRC 254e.

2) Forde, Alfred & Earl Sohm. United Nations System Cooperation in 1985. Developing Evaluation by Governments. Geneva: Joint Inspection Unit.

3) International Development Research Centre. 1986. With Our Own Hands. Research for Development: Canada's Contribution Through the IDRC 1970-1985. IDRC, Ottawa, Ontario, Canada. IDRC 246e.

4) Luukonen-Gronow, T. 1987. Scientific Research Evaluation: a review of methods and various contexts of their application. R&D Management, 17(3), 207-221.

5) World Bank/IMF. 1985. Report of the Task Force on Concessional Flows (Cassen Report). Development Committee. World Bank. Washington, D.C.

CATEGORIES AND METHODOLOGIES USED IN EVALUATION OF RESEARCH

1. **There are three main categories of evaluation:**

   i) ex ante or project appraisals refer to studies made before or shortly after the decision to support a project;

   ii) monitoring or mid-term review typically focus on operational activities such as project management and implementation. Some donors include monitoring in the evaluation process although others see it as essentially different;

   iii) ex/post evaluations and completion reports focus on lessons to be learned from experience and on improving future projects. These are more apt to address broader policy concerns as well as outcomes and impacts.

   In practice there are many variations on these and it is difficult to make precise distinctions among the numerous types used by donors. In one respect they comprise a continuous process which can be broken down into specific activities carried out at different points in the life of a project or program. As one moves through these, the concerns of the evaluator change. The distinction between the different types is important to avoid seeing evaluation, particularly ex-post studies, as simply an extension of monitoring or alternatively to think of it occuring only at the end of an activity.

2. **Purposes of evaluation of research projects:**

   i) quality of research and contribution to science;
   ii) project or program management;
   iii) development impact, effects/utilization of research results.

   Criteria of studies need to reflect the purposes for which evaluation is undertaken.

3. SUMMARY OF MAJOR EVALUATION METHODS AND THEIR APPLICATION

Major methods or techniques	Evaluated area	Criteria used	Phase of activity
Peer review	Disciplinary research areas: Appointments, project allocations, article screening; Research fields; Research institutes; Research programs	Internal	ex ante ongoing ex post
Modification of peer review: expert groups with peers & clients or potential users	Applied research activities: Research institutes; Research programs	Internal & external	ex ante ongoing
Interview and questionnaire methods	Applied research activities: Research institutes; Research programmes; Funding organizations	Internal & external	ex post
Quantitative methods: Methods for calculating profitability & social benefits	Applied R&D activities: Project assessment	External	ex ante
Technology indicators	R&D and innovation activities of industries and branches	External	ex post
Bibliometrics	Disciplinary research areas: University departments; Research fields; Research Institutes; Research groups	Internal	ex post
Case studies and histories	Innovations; R&D programmes	External	ex post

Source : T. Luukkonen-Gronow "Scientific Research Evaluation: a review of methods and various contexts of their application" R&D Management 17, 3, 1987 (pp. 209-221).

4. 'Internal' criteria arise from within the scientific professions and relate to the competence of the researcher and to the contribution of the activity to the advancement of science.

   The focus of 'internal' studies tends to be on technical questions at the frontiers of science or on management issues affecting the technical competence of the researchers.

5. 'External' criteria arise from outside the scientific disciplines and relate to socio-economic policy issues determined by the needs and circumstances of different countries where the research occurs.

DECISION-MAKERS/USERS	ISSUES	INDICATORS/CRITERIA
DONORS - public accountability - research for development - development projects **National Governments** - economic policy & strategic planners - science & related policy planners - development (social) policy **Research Institution** - managers - researchers **TIME FRAME** EX-ANTE MONITORING MID-TERM EX-POST	**RESEARCH PROCESS** - why do R&D projects succeed or fail - how to measure R&D productivity - how to measure effectiveness of funding mechanisms & training on R&D productivity - how to measure research institution capability - what kind of information (& support services) do scientists need to make a breakthrough **RESEARCH OUTCOME (DEVELOPMENT IMPACT)** - how to monitor research for its potential development impact - how to measure effectiveness of training - how to link researchers to users to improve transfer & utilization of research results. **STRATEGIC FORECASTING** - how to identify gaps where R&D resources could be most effectively used in context of national priorities and capabilities - how to measure impact of R&D expenditures on national economic development - how to measure receptivity of national environment to implement R&D results	**INPUT** [immediately measurable aspects that contribute to doing science] Resources - financial scientific personnel management personnel support services training/information Quality of researchers Contribution to scientific program Links to potential users Links to S&T infrastructure **OUTPUT** [measure direct results of R&D activities] publications prototypes and patents methodologies trained researchers industrial applications applications-oriented reports **IMPACT** [measure social & economic consequences of R&D on society] changes in productivity sales of high/improved technology products reduced imports of technology improvements in health & employment etc. **TARGETTING** [strategic analysis of where R&D resources could be most effectively allocated to contribute to development]